Sense Sentiment Similarity: An Analysis

Mitra Mohtarami et al. at AAAI 2012

Shih-Ming Wang

NLPLab, Institute of Information Science, Academia Sinica

February 13, 2016

Outline

Introduction

- Word similarity - which similarity?
- A brief History[1]
- Distributional Semantic - represent a word by its context
  - Document as context - LSA, LDA
    learns semantic relatedness (e.g. "boat" – "water")
  - Nearby words as context - *word2vec*, PMI factorization
    learns semantic similarity (e.g. "boat" – "ship")
- Not capturing accurate sentiment polarity?

[1] https://www.gavagai.se/blog/2015/09/30/a-brief-history-of-word-embeddings/

Topic Model

- Assume documents belong to some hidden topics, and each topic has different frequent words
- Latent Semantic Analysis (LSA) - apply SVD to factor term-document co-occurrence matrix ($M = W \Sigma D^T$)
- Latent Dirichlet Allocation (LDA) - Bayesian analysis on unigram model
  - Assume k topics, each represented by V dimension vector as distribution over vocabulary
  - Each word represented by k dimension vector as distribution over topics

Distributional Ccontext

- PMI factorization - apply SVD on word-word concurrence matrix ($M = W \Sigma C^T$)
- skip-gram - optimizing the probability of the concurrence of a word and its nearby words
- The two have been shown to be alike theoretically and empirically
- For word similarity, distributional context is better than topic models

Word Vector for Sentiment Similarity I

- Goal - measure sentiment similarity given two words X, Y
- Methods
  - Construct two d-dimensional vectors $\vec{X}, \vec{Y}$
  - Apply similarity function $Sim(X, Y)$, (cosine, correlation?)
  - Determine a threshold for Sim (middle of the range of S?)
- Solution
  - Measure the relation of a word to 12 emotion categories
  - Apply correlation

    $$Sim(X, Y) = \sum_{i=1}^{d} (\vec{X}_i - \bar{X})(\vec{Y}_i - \bar{Y})/(n-1)S_{\vec{X}}S_{\vec{Y}}$$
  - Determine a threshold by considering synonyms, antonyms of X, Y

Word Vector for Sentiment Similarity II

- **Step1** Build Vectors
- Emotion categories: anger, disgust, fear, guilt, sadness, shame, interest, joy, surprise, desire, love, courage
- Select synonyms for each category as seeds
  - Balance - Select equal amount of synonyms for each category
  - Relevant - Choose most similar synonyms according to semantic similarity scores computed by LSA
- For a word X and a category cat_k
  $$\vec{X}_k = \sum_{seed_i \in cat_k} coocur(X, seed_i)$$
- Problem: coocur is often 0
- Solution: $\vec{X}_k = \sum_{W \in synset(X)} \sum_{seed_i \in cat_k} coocur(W, seed_i)$

Word Vector for Sentiment Similarity III

- **Step2** Similarity Function $Sim(X, Y) = corr(\vec{X}, \vec{Y})$
- **Step3** Similarity Threshold
- For two similar words X, Y, and their antonyms ∼X, ∼Y
  - $Sim(\vec{X}, \vec{Y}) > Sim(\vec{X}, \sim \vec{Y})$
  - $Sim(\vec{X}, \vec{Y}) > Sim(\sim \vec{X}, \vec{Y})$
- Threshold: $Max\{Sim(\vec{X}, \sim \vec{Y}), Sim(\sim \vec{X}, \vec{Y})\}$
- With threshold 0, define
  $Sim(X, Y) = corr(\vec{X}, \vec{Y}) - Max\{corr(\vec{X}, \sim \vec{Y}), corr(\sim \vec{X}, \vec{Y})\}$
- Empirically better than taking $Sim(X, Y) = corr(\vec{X}, \vec{Y})$

Tasks

- Indirect yes/no question answer pairs (IQAP) Inference
  - $Sim(Adj_Q, Adj_A) > 0$ leads to answer yes
- Sentiment Orientation Prediction
  - Pick 7 pwords and 7 nwords
  - polarity(w) $= \sum_{p \in pwords} Sim(w, p) - \sum_{p \in nwords} Sim(w, n)$
- Compare different similarity function PMI, LSA, proposed

Settings

- Training
  - 50k movie reviews for calculating concurrence (PMI & proposed)
  - TASA, 61k documents for LSA
- Testing
  - IQAP, 125 Question/Answer pairs
  - MPQA 4000 positive/negative words

Results

Method	Precision	Recall	F1
PMI	60.61	58.70	59.64
LSA	66.70	54.95	60.26
Proposed	75.03	77.85	76.41

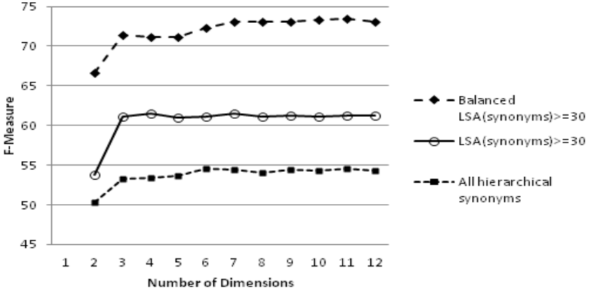
Table : IQAP Results

Method	Precision	Recall	F1
PMI	56.20	56.36	55.01
LSA	66.31	66.89	66.26
Proposed	73.07	73.89	73.11

Table : Sentiment Polarity Results

Error Analysis I

- Noise in emotion categories?
  - Apply SVD on word-category matrix ($|V|x12$)
  - 11 dimension achieves best F1 on sentiment polarity task
- Balance and relevance



Error Analysis II

- Role of synonyms and antonyms

Strategies	Precision	Recall	F1
w/o Ants and Syns	67.79	68.47	67.57
with Syns	71.47	72.25	71.43
with Ants	68.34	69.04	68.12
with Ants and Syns	**73.07**	**73.89**	**73.11**

Conclusion

- A method to construct word vector from prior knowledge is proposed
- Correlation is used to measure sentiment similarity for words
- Outperforms baselines on two tasks